THERAPEUTIC ADVANCES in
Medical Oncology

Original Research

Multivariable model for predicting 5-year survival in patients with EGFR-mutated non-small cell lung cancer treated with EGFR tyrosine kinase inhibitors: a retrospective study

Ther Adv Med Oncol

2025, Vol. 17: 1–16

DOI: 10.1177/17588359251321901

© The Author(s), 2025.
Article reuse guidelines: sagepub.com/journals-permissions

Qi-An Wang*, I-Lin Tsai*, Chien-Yu Lin, Po-Lan Su, Chien-Chung Lin, John Wen-Cheng Chang, Chen-Yang Huang, Yueh-Fu Fang, Ching-Fu Chang, Chih-Hsi Scott Kuo, Ping-Chih Hsu, Cheng-Ta Yang and Chiao-En Wu

Abstract

Background: Non-small-cell lung cancer (NSCLC) is the leading cause of cancer-related mortality worldwide. In Asian populations, epidermal growth factor receptor (EGFR) mutations are particularly prevalent, leading to the development of EGFR tyrosine kinase inhibitors (TKIs) to improve patient outcomes. While extensive research has been conducted on the prognosis of patients receiving EGFR-TKIs, the estimation of 5-year survival in this population remains an underexplored area.

Objectives: This study aimed to provide real-world evidence and conduct a comprehensive analysis of the determinants influencing the 5-year survival rate in patients with EGFR-mutated NSCLC. Considering the factors identified in this study, a scoring system was developed to predict the likelihood of patients achieving this goal.

Design: A retrospective cohort study utilizing a training cohort of 1,873 patients and a validation cohort of 484 patients.

Methods: A logistic regression model was constructed to evaluate the weighting of factors and develop a scoring system. The Kaplan-Meier model estimated the overall survival probability, and patients were categorized into four risk groups based on their likelihood of five-year survival. The prediction performance of both the training and validation cohorts was evaluated using the area under the curve (AUC), accuracy, precision, sensitivity, specificity, and F1-score.

Results: Results indicated that age > 65 years; performance score of 2-4; metastasis to the liver, brain, bone, or pleura; and poor disease control were associated with a decreased likelihood of 5-year survival. The estimated 5-year survival rate was 23.4% (odds ratio [OR]: 20.56; 95% confidence interval [CI]: 9.06-46.64; $p < 0.0001$), 16.1% (OR: 12.88; 95% CI: 5.82-28.49; $p < 0.0001$), 7.2% (OR: 5.23; 95% CI: 2.36-11.60; $p < 0.0001$), and 1.5% (OR: reference) for the low-risk, intermediate-risk, high-risk, and very-high-risk groups, respectively. The validation cohort further confirmed these findings, showing survival probabilities of 52.6% (OR: 96.67; 95% CI: 11.07-844.23; $p < 0.0001$), 21.3% (OR: 23.49; 95% CI: 3.13-176.46; $p = 0.002$), 14.9% (OR: 15.21; 95% CI: 2.03-114.25; $p = 0.008$), and 1.1% (OR: reference) for the low-risk, intermediate-risk, high-risk, and very-high-risk groups, respectively. The training cohort demonstrated an AUC of 0.79 (95% CI: 0.75-0.82) and a model quality score of 0.75, indicating good predictive performance. Calibration plots demonstrated a good fit for the scoring system. For the external validation cohort, the AUC, precision, sensitivity, and specificity were 0.71, 0.74, 0.35, 0.33, respectively. The model achieved an F1-score of 0.47, reflecting adequate performance in predicting 5-year survival probabilities.

Correspondence to:
Chiao-En Wu
Division of Hematology-Oncology, Department of Internal Medicine, New Taipei Municipal TuCheng Hospital, New Taipei City 236043, Taiwan

Division of Hematology-Oncology, Department of Internal Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, Taoyuan, Taiwan
jiaoen@gmail.com

Qi-An Wang
School of Medicine, College of Medicine, Chang Gung University, Taoyuan, Taiwan

I-Lin Tsai
Chien-Yu Lin
Po-Lan Su
Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Chien-Chung Lin
Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Institute of Clinical Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Department of Biochemistry and Molecular Biology, College of Medicine, National Cheng Kung University, Tainan, Taiwan

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the Sage and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

John Wen-Cheng Chang
Chen-Yang Huang
Ching-Fu Chang
Division of Hematology-Oncology, Department of Internal Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, Taoyuan, Taiwan

Yueh-Fu Fang
Chih-Hsi Scott Kuo
Ping-Chih Hsu
Cheng-Ta Yang
Division of Thoracic Oncology, Department of Thoracic Medicine, Linkou Chang Gung Memorial Hospital, College of Medicine, Chang Gung University, Taoyuan, Taiwan

*These authors are contributed equally to the article and are acknowledged as first co-authors.

Conclusion: This study identified critical prognostic factors and developed a validated scoring system for estimating 5-year survival in patients with EGFR-mutated NSCLC receiving EGFR-TKIs. While the model demonstrated robust predictive performance within the study cohort, broader applicability beyond Taiwan may require further refinements and alternative study designs.

Keywords: 5-year survival, EGFR mutation, EGFR tyrosine kinase inhibitors, non-small-cell lung cancer

Received: 17 October 2024; revised manuscript accepted: 4 February 2025.

Introduction

Lung cancer is among the primary global causes of mortality,[1] with approximately 80% of cases attributed to non-small-cell lung cancer (NSCLC), and 57% of patients with NSCLC presenting with metastasis.[2] For patients diagnosed with early-stage NSCLC, surgical resection has been the preferred treatment.[3] However, in the past, treatment options for those with advanced NSCLC were primarily limited to chemotherapy, which offered a progression-free survival (PFS) of only 5–6 months,[4,5] with approximately 4.5%–10% of the patients achieving 5-year survival.[1,6,7] Since the groundbreaking work of Todaor et al. in 1980,[8] which established the connection between NSCLC and epidermal growth factor receptor (EGFR), extensive research has focused on exploring the molecular aspects of this disease, significantly advancing the ability to combat this form of cancer. EGFR mutations are well-known drivers of NSCLC, occurring in 19% of Western patients, and a notably higher occurrence of approximately 50% in Asian patients.[9,10] Clinical trials of EGFR tyrosine kinase inhibitors (TKIs) have demonstrated their superior efficacy in improving PFS and overall survival (OS) in patients with EGFR-mutated NSCLC.[11–13] Therefore, EGFR-TKIs are now the standard treatment for EGFR-mutated NSCLC.[14] Despite extensive research on PFS and OS, relatively few studies have analyzed the analysis of 5-year survival in this population.

Therefore, this retrospective study aimed to provide real-world evidence and conduct a comprehensive analysis of the determinants influencing the 5-year survival rate in patients with EGFR-mutated NSCLC. Considering the factors identified in this study, a scoring system was developed to predict the likelihood of patients achieving this goal.

Methods

Study design and participants

The training cohort of this retrospective study analyzed data extracted from the Chang Gung Research Database (CGRD) of Cheng Gung Memorial Hospital (CGMH), a multi-institutional medical record registry with coverage of approximately 10% of the Taiwanese population.[15] Patients who met the following criteria were included: (1) pathologically confirmed diagnosis of NSCLC, (2) detection of EGFR mutation (exon 19 deletion or exon 21 L858R point mutation) in NSCLC, (3) treatment with EGFR-TKIs (afatinib, erlotinib, and gefitinib) as the first-line therapy without combination with other systemic treatments, and (4) age ⩾18 years. Ultimately, 1873 patients were enrolled between 2011 and 2016, ensuring at least 5 years of follow-up until the last follow-up in 2021 or mortality (Figure 1). For the validation cohort, the dataset was obtained from the National Cheng Kung University Hospital (NCKUH), a tertiary medical center in southern Taiwan. Identical inclusion criteria were applied to the validation cohort to further evaluate the utility of the scoring system. Five years of follow-up or mortality was ensured with the last follow-up in June 2022, resulting in 484 patients available for analysis (Figure 2).

De-identified patient data were collected, including age, sex, performance score, smoking history, morphology of NSCLC, genotypes of EGFR mutations, staging of NSCLC, EGFR-TKIs used, metastatic status (liver, brain, bone, pleura, adrenal, distant lymph node, pericardia, and peritoneal), objective response rate, and disease control rate. This retrospective study was approved by the Institutional Review Boards of CGMH (IRB: 202401619B0) and NCKUH (IRB: B-ER-112-118). The requirement for informed consent was waived owing to the retrospective nature of this study. The reporting of

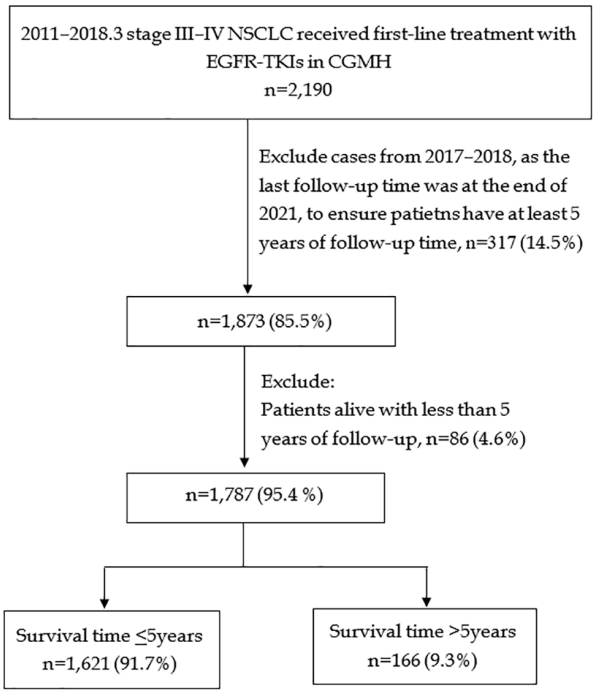

Figure 1. Study flow diagram of the CGMH (training cohort).
CGMH, Cheng Gung Memorial Hospital.

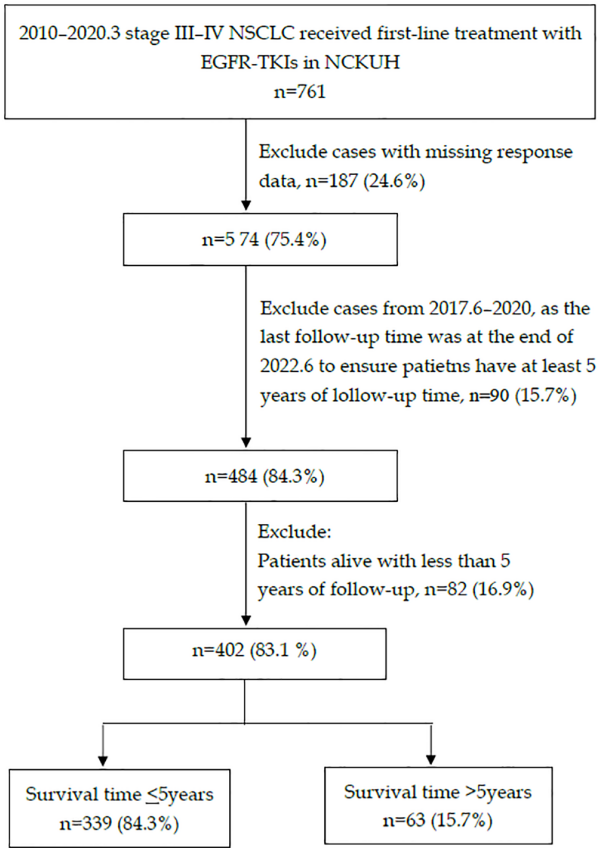

Figure 2. Study flow diagram of the NCKUH (validation cohort).
NCKUH, National Cheng Kung University Hospital.

this study conformed to the STROBE guideline (Supplemental Table 1).

Statistical analysis

Patient demographic data were obtained from the collected information. The odds ratios (ORs) were calculated using univariate and multivariate analyses. Variables with statistical significance ($p < 0.05$) in the multivariate analysis were further selected to develop a scoring system based on a logistic regression model to evaluate the weight of factors on patients' 5-year survival. The OS probability was estimated using the Kaplan–Meier model, and patients were categorized into four risk groups (low, intermediate, high, and very high) based on their chances of surviving beyond 5 years. Hazard ratios (HRs) were calculated at different time points (up to, and after 5 years). The accuracy of the scoring system was assessed using the area under the receiver operating characteristic curve (AUC) and precision-recall curve, whereas calibration was evaluated using a calibration plot. Furthermore, the prediction performance of the validation cohort was evaluated using the AUC, accuracy, precision, sensitivity, specificity, and F1-score. Statistical analyses, including baseline patient characteristics and AUC, were performed using Statistical Package for Social Sciences (Version 26.0; IBM SPSS Statistics for Windows, IBM Corp. Armonk, NY, USA). In addition, ORs from univariate and multivariate analyses, the scoring system based on the logistic regression model, the Kaplan–Meier analysis, validation of the prediction model, and survival curve analyses were conducted using the R Statistical Software (Version 4.4.1; R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

Results

Patient characteristics (baseline and those who survived for >5 years)

Of the 1873 enrolled patients, 86 were lost to follow-up within 5 years, leaving 1787 for analysis. The baseline characteristics of the patients are summarized in Table 1. Among the patients, 992 (55.5%) were aged >65 years, and 1068 (59.8%) were female. A total of 1384 (77.4%) patients had a performance score of 0–1, whereas 1307 (73.1%) were non-smokers. Regarding pathology, 1754 (98.2%) patients were confirmed as adenocarcinoma, 944 (52.8%) had exon 21 L858R point mutation, and 1674 (93.7%) had stage IV NSCLC. For treatment, 1001 (56%), 416 (23.3%), and 370 (20.7%) patients received gefitinib, afatinib, and erlotinib, respectively. Metastatic status varied widely, with the highest number of patients experiencing bone metastasis (841 patients, 47.1%) and the lowest number having adrenal metastasis (163 patients, 9.1%). Regarding prognosis after EGFR-TKI treatment, 1440 (80.6%) patients achieved complete response (CR), partial response (PR), or stable disease (SD) for their disease control rate, and 166 (9.3%) survived for more than 5 years.

Factors associated with prolonged survival

Among the 166 patients who achieved 5-year survival with $p < 0.05$ (Table 2), 97 (58.4%) were aged <65 years; 156 (94%) had a performance score of 0–1; 134 (80.7%) had stage IV disease; 4 (204%), 33 (19.9%), 39 (23.5%), 48 (28.9%), and 8 (4.8%) had no metastasis of the liver, brain, bone, pleura, and adrenal gland; and 160 (96.4%) had an adequate disease control rate. There were no significant differences between the patients' sex, smoking history, morphology of NSCLC, EGFR mutation genotypes, applied EGFR-TKIs, or metastatic status of the lung.

In addition, variables with $p < 0.05$ from Table 2 were used to calculate the ORs of patients surviving for more than 5 years. The subgroups with a lower population in each category were set as a reference, and univariate and multivariate analyses were applied (Table 3). Although cancer staging reached statistical significance ($p < 0.0001$), as shown in Table 2, it was not considered an independent variable and may interact with other factors, such as metastasis to different organs. Thus, it is excluded from Table 3.

In the univariate analysis, age of ⩽65 years (OR: 1.86; 95% confidence interval (CI): 1.35–2.57; $p < 0.001$) and a performance score of 0–1 (OR: 4.99; 95% CI: 2.61–9.56; $p < 0.0001$) were associated with the increased probability of surviving for more than 5 years. Moreover, the absence of metastasis in the liver (OR: 6.97; 95% CI: 2.56–18.97; $p < 0.001$), brain (OR: 1.71; 95% CI: 1.15–2.54; $p = 0.008$), bone (OR: 3.19; 95% CI: 2.20–4.63; $p < 0.0001$), pleura (OR: 2.21; 95% CI: 1.56–3.14; $p < 0.0001$), and adrenal gland (OR: 2.09; 95% CI: 1.01–4.33; $p = 0.048$) was also linked to a likelihood of achieving 5-year survival. For disease control rate, CR, PR, and SD

Table 1. Patients' baseline characteristics ($n = 1787$).

Characteristics	No. of patients	%
Age (years)		
<65	795	44.5
>65	992	55.5
Sex		
Male	719	40.2
Female	1068	59.8
Performance score		
0–1	1384	77.4
2–4	403	22.6
Smoking		
No	1307	73.1
Yes	434	24.3
Unknown	46	2.6
Morphology		
Adenocarcinoma	1754	98.2
Non-adenocarcinoma	33	1.8
Mutation		
Exon 19 deletion	843	47.2
L858R	944	52.8
Stage		
IIIB	113	6.3
IV	1674	93.7
TKI drug		
Afatinib	416	23.3
Erlotinib	370	20.7
Gefitinib	1001	56.0
Liver metastasis	242	13.5
Brain metastasis	516	28.9
Bone metastasis	841	47.1
Pleura metastasis	816	45.7
Adrenal metastasis	163	9.1
Disease control rate		
CR/PR/SD	1440	80.6
PD/NA	347	19.4
Survival time (years)		
<5	1621	90.7
>5	166	9.3

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NA, not assessed; TKI, tyrosine kinase inhibitor.

(OR: 7.10; 95% CI: 3.12–16.19; $p < 0.0001$) were correlated with an increased probability of surviving for 5 years.

Considering these results, a multivariate analysis was performed. Age ⩽65 years (OR: 2.05; 95% CI: 1.45–2.39; $p < 0.0001$), a performance score of 0–1 (OR: 3.09; 95% CI: 1.58–6.03; $p < 0.001$), and CR, PR, and SD disease control rate (OR: 5.53; 95% CI: 2.40–12.75; $p < 0.0001$) remained significantly correlated with an increased probability of surviving for more than 5 years. In addition, the absence of metastasis to the liver (OR: 5.04; 95% CI: 1.82–13.97; $p = 0.002$), brain (OR: 1.67; 95% CI: 1.09–2.58; $p = 0.02$), bone (OR: 3.32; 95% CI: 2.23–4.92; $p < 0.0001$), and pleura (OR: 2.75; 95% CI: 1.90–4.00; $p < 0.0001$) were also associated with higher chances of 5-year survival. However, the absence of metastasis to the adrenal gland (OR: 1.47; 95% CI: 0.68–3.17; $p = 0.328$) did not exhibit statistical significance, unlike in the univariate analysis.

Scoring system

A scoring system (Table 4) was developed using a logistic regression model, in which points were assigned to each prognostic factor associated with an elevated risk of not achieving 5-year survival. The factors identified as statistically significant variables in the multivariate analysis ($p < 0.05$) were as follows: age >65 years (1 point); performance score of 2–4 (2 points); metastasis to the liver (3 points), brain (1 point), bone (1 point), or pleura (2 points); and poor disease control (3 points).

Table 2. Characteristics of patients who achieved and not achieved 5-year survival.

Variables	Survival time		*p* Value
	<5 years (*n* = 1621)	>5 years (*n* = 166)	
Age (years)			<0.001
<65	698 (43.1%)	97 (58.4%)	
>65	923 (56.9%)	69 (41.6%)	
Sex			0.050
Male	664 (41.0%)	55 (33.1%)	
Female	957 (59.0%)	111 (66.9%)	
Performance score			<0.0001
0–1	1228 (75.8%)	156 (94.0%)	
2–4	393 (24.2%)	10 (6.0%)	
Smoking			0.362
No	1178 (72.7%)	129 (77.7%)	
Yes	400 (24.7%)	34 (20.5%)	
Unknown	43 (2.7%)	3 (1.8%)	
Morphology			0.763
Adenocarcinoma	1590 (98.1%)	164 (98.8%)	
Non-adenocarcinoma	31 (1.9%)	2 (1.2%)	
Mutation			0.353
Exon 19 deletion	759 (46.8%)	84 (50.6%)	
L858R	862 (53.2%)	82 (49.4%)	
Stage			<0.0001
IIIB	81 (5.0%)	32 (19.3%)	
IV	1540 (95.0%)	134 (80.7%)	
TKI drug			0.312
Afatinib	371 (22.9%)	45 (27.1%)	
Erlotinib	333 (20.5%)	37 (22.3%)	
Gefitinib	917 (56.6%)	84 (50.6%)	
Liver metastasis	238 (14.7%)	4 (2.4%)	<0.0001
Brain metastasis	483 (29.8%)	33 (19.9%)	0.007
Bone metastasis	802 (49.5%)	39 (23.5%)	<0.0001
Pleura metastasis	768 (47.4%)	48 (28.9%)	<0.0001
Adrenal metastasis	155 (9.6%)	8 (4.8%)	0.043
Disease control rate			<0.0001
CR/PR/SD	1280 (79.0%)	160 (96.4%)	
PD/NA	341 (21.0%)	6 (3.6%)	

CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NA, not assessed; TKI, tyrosine kinase inhibitor.

Table 3. Univariate and multivariate analyses (ORs for surviving over 5 years).

Variables	Univariate analysis		Multivariate analysis	
	OR	*p* Value	Adjust OR	*p* Value
Age				
<65 vs >65	1.86 (1.35–2.57)	<0.001	2.05 (1.45–2.39)	<0.0001
Performance score				
0–1/2–4	4.99 (2.61–9.56)	<0.0001	3.09 (1.58–6.03)	<0.001
Liver metastasis				
No/yes	6.97 (2.56–18.97)	<0.001	5.04 (1.82–13.97)	0.002
Brain metastasis				
No/yes	1.71 (1.15–2.54)	0.008	1.67 (1.09–2.58)	0.020
Bone metastasis				
No/yes	3.19 (2.20–4.63)	<0.0001	3.32 (2.23–4.92)	<0.0001
Pleura metastasis				
No/yes	2.21 (1.56–3.14)	<0.0001	2.75 (1.90–4.00)	<0.0001
Adrenal metastasis				
No/yes	2.09 (1.01–4.33)	0.048	1.47 (0.68–3.17)	0.328
Disease control rate				
CR/PR/SD	7.10 (3.12–16.19)	<0.0001	5.53 (2.40–12.75)	<0.0001
PD/NA	Reference		Reference	

CR, complete response; OR, odds ratio; PR, partial response; SD, stable disease; PD, progressive disease; NA, not assessed.

Furthermore, the Kaplan–Meier model was used to demonstrate the OS probability of different risk groups achieving 5-year survival. Patients were categorized based on observed 5-year survival rates, with clinical significance guiding the grouping: low risk (>20%), intermediate risk (10%–20%), high risk (5%–10%), and very high risk (<5%), with corresponding scores assigned to reflect increasing risk levels (Table 5 and Figure 3). The estimated 5-year survival rate was 23.4% ($n = 44$; OR: 20.56; 95% CI: 9.06–46.64; $p < 0.0001$), 16.1% ($n = 62$; OR: 12.88; 95% CI: 5.82–28.49; $p < 0.0001$), 7.2% ($n = 53$; OR: 5.23; 95% CI: 2.36–11.60; $p < 0.0001$), and 1.5% ($n = 7$; OR: reference) for the low-risk (score 0), intermediate-risk (score 1–2), high-risk (score 3–5), and very-high-risk (score 6–14) groups, respectively. The number of survivors at each time point is displayed in Figure 3. The majority of patients were categorized into the high-risk group ($n = 735$), whereas the low-risk group had the smallest number of patients ($n = 188$). Patients in the high-risk and very-high-risk groups all died by the 108th month, with two patients in the low-risk group and one patient in the intermediate-risk group still alive. However, none of the patients survived for 10 years. The median OS was 43.7 months (95% CI: 36.6–50.7) for the low-risk group, 26.5 months (95% CI: 24.0–29.0) for the intermediate-risk group, 21.2 months (95% CI: 19.5–22.8) for the high-risk group, and 10.9 months (95% CI: 9.8–12.0) for the very-high-risk group.

In addition, HRs at different time points were calculated using the low-risk group as the reference

Table 4. Scoring system to calculate point values for risk score.

Variables	β[a]	Categories	Reference value (W)	$\beta (W - W_{REF})$	Points = $\beta (W - W_{REF})/B$[b]
Age	0.713	<65[a]	0 (W_{REF})	0	0
		>65	1	0.713	1
Performance score	1.131	0–1[a]	0 (W_{REF})	0	0
		2–4	1	1.131	2
Liver metastasis	1.647	Yes	1	1.647	3
		No[a]	0 (W_{REF})	0	0
Brain metastasis	0.530	Yes	1	0.530	1
		No[a]	0 (W_{REF})	0	0
Bone metastasis	1.212	Yes	1	1.212	2
		No[a]	0 (W_{REF})	0	0
Pleura metastasis	1.004	Yes	1	1.004	2
		No[a]	0 (W_{REF})	0	0
Disease control rate	1.706	CR/PR/SD[a]	0 (W_{REF})	0	0
		PD/NA	1	1.706	3

[a]Regression coefficients (β) of the multivariate logistic regression model.
[b]Base constant (constant B) using the smallest regression coefficient in the model.
CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; NA, not assessed.

Table 5. Corresponded probability of achieving 5-year survival to each risk group.

Risk group	Score	*N*	Survived >5 years, No. (%)	Odds ratio (CI)	*p* Value
Low	0	188	44 (23.4%)	20.56 (9.06–46.64)	<0.0001
Intermediate	1–2	386	62 (16.1%)	12.88 (5.82–28.49)	<0.0001
High	3–5	735	53 (7.2%)	5.23 (2.36–11.60)	<0.0001
Very high	6–14	478	7 (1.5%)	Reference	

Low-risk group (score 0–1): >20% reaching 5-year survival. Intermediate-risk group (score 2): 10%–20% reaching 5-year survival. High-risk group (score 3): 5%–10% reaching 5-year survival. Very-high-risk group (score 4–10): <5% reaching 5-year survival.
CI, confidence interval.

(Table 6). Before the fifth year, compared to the low-risk group, the intermediate-risk group had a 30% higher likelihood of not achieving 5-year survival (HR: 1.30; 95% CI: 1.06–1.58; $p = 0.01$), the high-risk group had a 95% higher likelihood (HR: 1.95; 95% CI: 1.63–2.34; $p < 0.0001$), and the very-high-risk group had a 4.27 times higher likelihood (HR: 4.27; 95% CI: 3.54–5.17; $p < 0.0001$). After the fifth year, the intermediate-risk group had a 15% higher likelihood of death (HR: 1.15; 95% CI: 0.56–2.35; $p = 0.702$), the high-risk group had a 2.10 times higher likelihood (HR: 2.10; 95% CI: 1.05–4.17; $p = 0.035$), and the very-high-risk group had a 2.34 times higher likelihood (HR: 2.34; 95% CI: 0.76–7.21; $p = 0.139$). The overall HRs were 1.28 (95% CI:

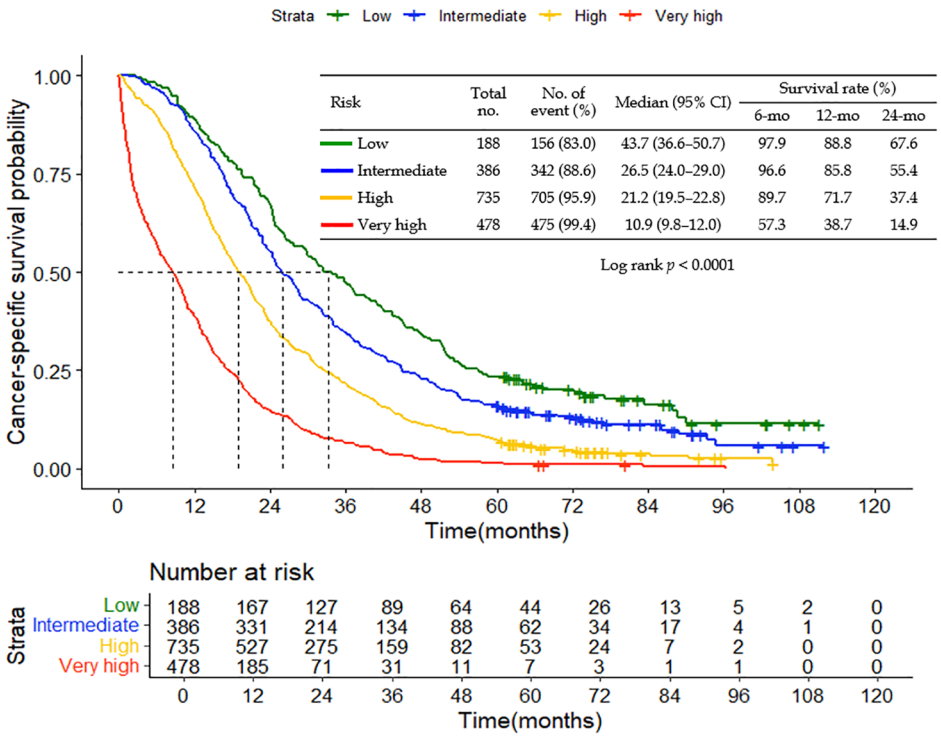

Figure 3. Kaplan–Meier model demonstrating the overall survival probability.

Table 6. HRs at different time points.

Risk group	Low ($n=188$)	Intermediate ($n=386$)	High ($n=735$)	Very high ($n=478$)
HR	Reference	1.28	1.95	4.22
95% CI		1.06–1.55	1.64–2.32	3.51–5.08
p value		0.010	<0.0001	<0.0001
HR up to 5 years	Reference	1.30	1.95	4.27
95% CI		1.06–1.58	1.63–2.34	3.54–5.17
p value	—	0.010	<0.0001	<0.0001
HR after 5 years	Reference	1.15	2.10	2.34
95% CI		0.56–2.35	1.05–4.17	0.76–7.21
p value	—	0.702	0.035	0.139

CI, confidence interval; HR, hazard ratio.

1.06–1.55; $p=0.01$) for the intermediate-risk group, 1.95 (95% CI: 1.64–2.32; $p<0.0001$) for the high-risk group, and 4.22 (95% CI: 3.51–5.08; $p<0.0001$) for the very-high-risk group.

To further assess the accuracy of the proposed model in predicting the likelihood of patients reaching 5-year survival, AUC and PR curves were obtained (Figure 4). The calculated AUC (area: 0.79; 95% CI: 0.75–0.82) indicated an adequate result, with an overall model quality of 0.75. For the calibration of the scoring system, a calibration plot between the actual and predicted probabilities of reaching 5-year survival (Figure 5) was obtained, demonstrating a good model fit.

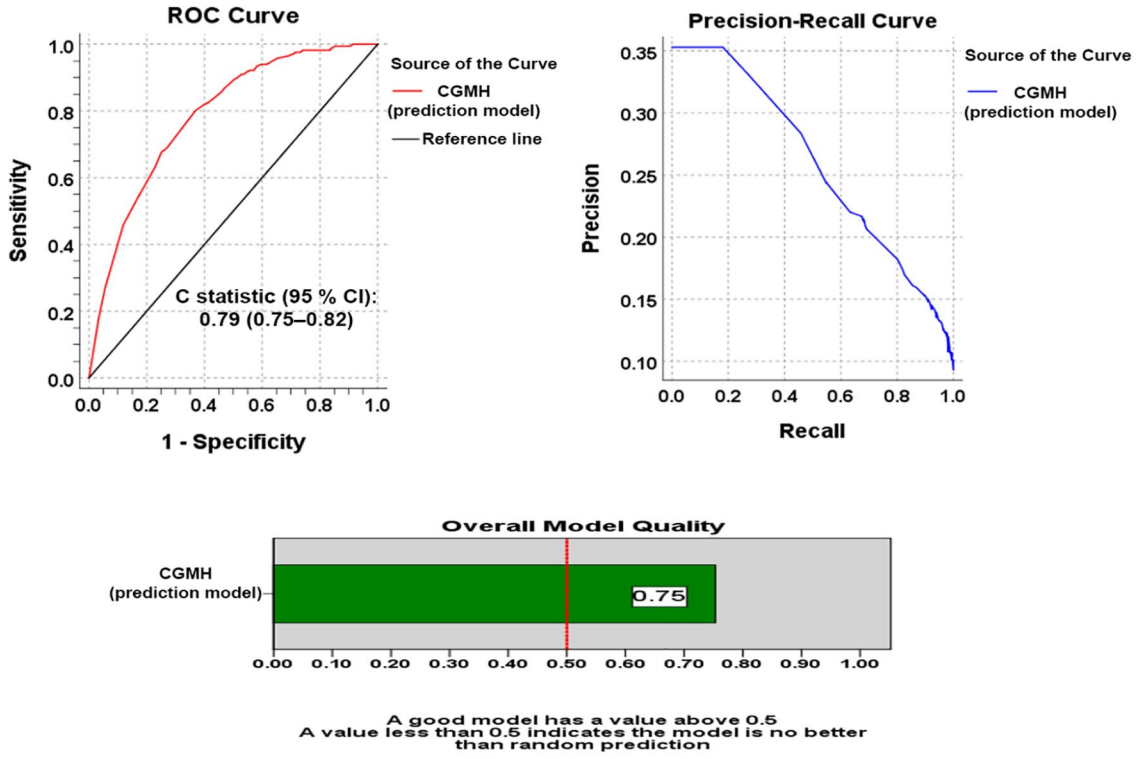

Figure 4. Receiver operating characteristic curve and precision-recall curve for the prediction model.

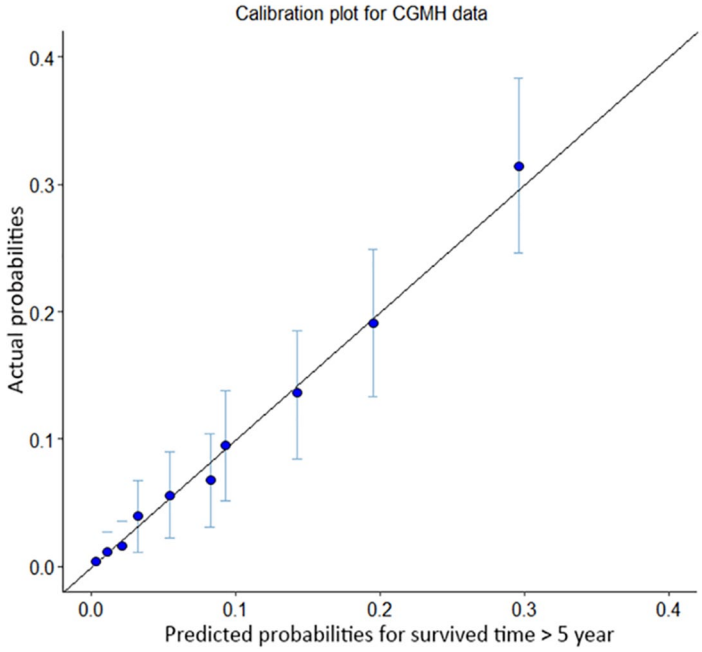

Figure 5. Calibration plot between the actual and predicted probability of reaching 5-year survival.

Verification of the scoring system by the second set of data

The training cohort was validated using a dataset collected from the NCKUH (Table 7). Of the 484 enrolled patients, 82 were lost to follow-up, resulting in a final sample size of 402 patients. A comparison of the baseline characteristics of the training and validation cohorts is shown in

Table 7. Probability of achieving 5-year survival in each risk group (validation cohort).

Risk group	Score	*N*	Survived >5 years, No. (%)	Odds ratio (CI)	*p* Value
Low	0	19	10 (52.6)	96.67 (11.07–844.23)	<0.0001
Intermediate	1–2	127	27 (21.3)	23.49 (3.13–176.46)	0.002
High	3–5	168	25 (14.9)	15.21 (2.03–114.25)	0.008
Very high	6–10	88	1 (1.1)	Reference	

CI, confidence interval.

Supplemental Table 2. The low-risk group comprised the fewest individuals ($n=19$), with a survival probability of 52.6% ($n=10$; OR: 96.67; 95% CI: 11.07–844.23; $p<0.0001$). The survival probability decreased to 21.3% ($n=27$; OR: 23.49; 95% CI: 3.13–176.46; $p=0.002$) in the intermediate-risk group and 14.9% ($n=25$; OR: 15.21; 95% CI: 2.03–114.25; $p=0.008$) in the high-risk group. The very-high-risk group had the lowest survival probability of 1.1% ($n=1$; OR: reference). All patients in the very-high-risk group died by the 96th month, whereas none of the patients in the high-risk group survived beyond the 108th month. Four patients achieved 10-year survival (one from the low-risk group and three from the intermediate-risk group); however, none survived past the 132nd month. These results corroborate our previous findings from the CGRD dataset, confirming that the low-risk group had the highest probability of 5-year survival, with the survival probability decreasing as the risk scores increased.

The prediction performance results of the external validation are summarized in Figure 6. The calculated AUC was 0.71 with a precision, sensitivity, and specificity of 0.74, 0.35, and 0.33, respectively. The F1-score for the model was 0.47, indicating adequate performance in predicting the probability of reaching 5-year survival for these patients.

Discussion

In our patient population from the CGRD, a total of 166 patients (9.3%) achieved 5-year survival, which is consistent with existing studies. The U.S. Surveillance, Epidemiology, and End Results program reported 5-year survival rates of 11.0% and 6.9% for patients with adenocarcinoma and squamous cell carcinoma, respectively, from 2013 to 2019.[16] In another study focusing on Asian patients, Jeon et al. reported 5-year survival rates of 16% and 10% for stage III and IV NSCLC, respectively; however, patients from both survivor and non-survivor groups were treated with various modalities, including surgery and radiotherapy.[6] By contrast, for a population similar to ours, Lin et al. conducted a study on 137 patients with EGFR-mutated metastatic lung carcinoma treated with erlotinib or gefitinib.[17] They reported an outstanding 5-year survival rate of 14.6%.[17]

For factors influencing the probability of achieving 5-year survival, we proposed that age, performance score, stage, metastatic status (liver, bone, and pleura), and disease control rate were associated with prolonged survival. However, Lin et al. suggested that exon 19 deletion, absence of extrathoracic or brain metastasis, and non-smoking status were associated with prolonged OS.[17] They also indicated that age; sex; stage at diagnosis; liver, bone, or adrenal metastasis; EGFR-TKIs; and the line of TKI therapy were not correlated with OS.[17] The importance of smoking status was emphasized in another Japanese study, along with sex, age, performance score, and clinical stage as independent prognostic factors.[18] In addition, Campos-Balea et al. reported that poor prognosis among patients with metastatic lung adenocarcinoma was indicated by male sex, age ⩾65 years old, and the presence of liver, bone, or brain metastasis.[19] Regarding genetic mutation, common EGFR mutations have been shown to display better clinical outcomes than uncommon EGFR mutations, such as L861Q or wild-type EGFR.[20–22] Furthermore, even among common EGFR mutations, such as exon 19 deletion or L858R, outcomes may vary. In the present study, genetic mutation (exon 19 deletion and exon 21 L858R) was not considered a prognostic factor

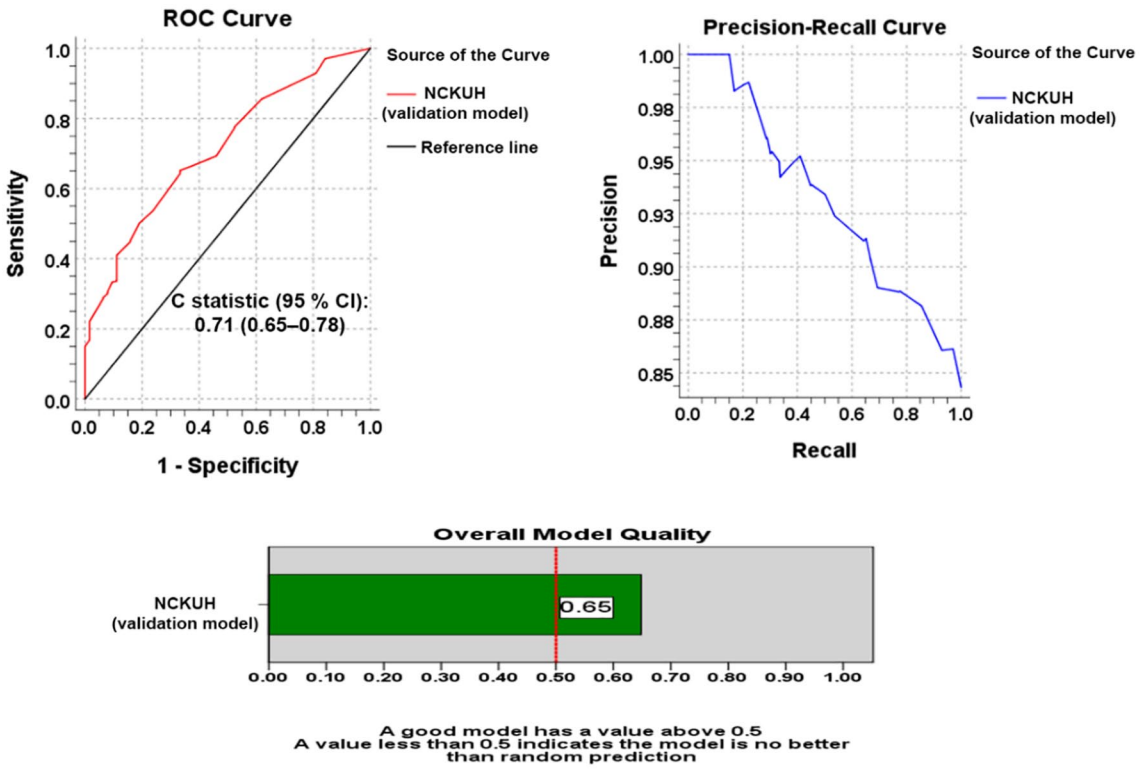

Figure 6. Receiver operating characteristic curve and precision-recall curve for the validation model.

for predicting the probability of reaching 5-year survival, which is not entirely consistent with the existing literature. Kuan et al. reported that patients with exon 19 deletions received a significant OS benefit from EGFR-TKI treatment, whereas those with L858R showed no OS benefit.[23] This result was subsequently confirmed by Leduc et al., who found that patients with exon 19 deletion consistently exhibited superior outcomes than individuals harboring L858R mutation.[22] The inconsistency between our findings and existing literature on the prognostic value of genetic mutations may be attributed to the selection of EGFR-TKIs. Although some studies have demonstrated that exon 19 deletion may be more sensitive to EGFR-TKI treatment than exon 21 L858R,[24] other studies have suggested that there was no statistical significance in terms of PFS between patients with the two genetic mutations receiving first-generation EGFR-TKIs (erlotinib and gefitinib).[25–27] In the study by Maemondo et al., under the treatment of gefitinib, the median PFS for exon 19 deletion and exon 21 L858R were 11.5 and 10.8 months, respectively.[25] Yang et al. conducted another study to compare the efficacy of gefitinib and erlotinib. In the subgroup analysis, the median PFS was 11.4 months for exon 19 deletion and 11.2 months for exon 21 L858R mutation. Afatinib has shown superior efficacy compared with erlotinib and gefitinib.[28,29] In terms of efficacy in different genetic mutations, exon 19 deletion was more favorable than exon 21 L858R as shown in the preplanned stratified analysis in both the LUX-Lung 3[30] and LUX-Lung 6[13] trials. However, as shown in Table 1, the distribution between each EGFR-TKI group was uneven, with gefitinib comprising most of the patients ($n = 1001$, 56%), followed by afatinib ($n = 416$, 23.3%) and erlotinib ($n = 370$, 20.7%). This uneven distribution may have contributed to the lack of inclusion of genetic mutations as a prognostic factor for predicting the possibility of reaching 5-year survival.

Beyond the first- and second-generation EGFR-TKIs, osimertinib has demonstrated superior PFS (median PFS: 18.9 months vs 10.2 months)[31] and OS (median OS: 38.6 months vs 31.8 months)[32] compared with erlotinib or gefitinib, as shown in the FLAURA study. Osimertinib was approved by the U.S. Food and Drug Administration in 2018, with subsequent approvals in other countries after the study period (2011–2016) for the training cohort (with

the latest follow-up in 2021). Since osimertinib was approved in Taiwan in 2020, it could not be analyzed in this cohort to ensure a 5-year follow-up period. Two phase 3 studies, FLAURA2 (osimertinib + chemotherapy, median PFS 25.2 months vs 16.7 months) and MARIPOSA (amivantamab + lazertinib, median PFS 23.7 months vs 16.6 months), have further demonstrated the potential to extend the superior PFS achieved with osimertinib. However, the benefits of these combinations on OS remain unclear. These treatment options have become the new standard of care for patients with EGFR-mutated NSCLC. However, biomarkers are required to stratify patients and guide them toward the most appropriate treatment. This study provides a potential model to address this issue. For patients expected to have favorable outcomes, EGFR-TKI monotherapy may suffice. By contrast, for patients predicted to have unfavorable outcomes, combination treatments such as the FLAURA2 or MARIPOSA regimens should be considered.

To the best of our knowledge, this is the first study to focus on the development of a scoring system for estimating the attainment of 5-year survival. Although some of the identified factors were inconsistent with previous literature, the AUC and calibration plot provided evidence supporting the accuracy of the system. However, further studies are required to examine the practicality of this method across different patient populations.

The limitations of our study are as follows: (1) The retrospective nature of our study may introduce undefined bias, potentially impacting the results. (2) Although our study used a multi-institutional design with coverage of nearly 10% of the Taiwanese population, it may not be representative of the entire Asian or Chinese community. (3) Osimertinib and other new treatments were not included in this study because of insufficient follow-up data to demonstrate 5-year efficacy. Therefore, the clinical significance of this model is limited to the era of third-generation EGFR-TKIs. However, many people in Asian countries still lack access to osimertinib as a first-line treatment,[33] making this model valuable for such populations.

Conclusion

The key determinants of achieving 5-year survival in patients with EGFR-mutated NSCLC treated with EGFR-TKIs were identified. This analysis, encompassing data from 1787 patients over a 5-year period, highlighted factors such as age, performance score, disease stage, metastatic status (specifically liver, bone, and pleura), and the disease control rate. Leveraging these findings, we introduced a scoring system tailored to predict 5-year survival in this specific patient group, which was further verified using a second dataset obtained from another tertiary medical center in Taiwan. Refining this prediction model for broader applications beyond Taiwan may necessitate alternative study designs.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Institutional Review Board of Chang Gung Memorial Hospital (IRB: 202401619B0) and National Cheng Kung University Hospital (IRB: B-ER-112-118).

Consent for publication

The requirement for informed consent was waived owing to the retrospective nature of this study by the Institutional Review Boards of the relevant institutions.

Author contributions

Qi-An Wang: Formal analysis; Investigation; Project administration; Validation; Writing – original draft; Writing – review & editing.

I-Lin Tsai: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Chien-Yu Lin: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Po-Lan Su: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Chien-Chung Lin: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualization; Writing – review & editing.

John Wen-Cheng Chang: Conceptualization; Data curation; Formal analysis; Investigation;

Methodology; Resources; Validation; Writing – review & editing.

Chen-Yang Huang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Yueh-Fu Fang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Ching-Fu Chang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Chih-Hsi Scott Kuo: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Ping-Chih Hsu: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Cheng-Ta Yang: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Resources; Validation; Writing – review & editing.

Chiao-En Wu: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing – review & editing.

Acknowledgements
None.

Funding
The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from Linkou Chang Gung Memorial Hospital (CMRPG3M0971 to C.-E.W.).

Competing interests

The authors declare that there is no conflict of interest.

Availability of data and materials
All data generated or analyzed during the study are included in this article. Further inquiries can be directed to the corresponding author.

ORCID iDs
Qi-An Wang https://orcid.org/0009-0008-7340-3405

I-Lin Tsai https://orcid.org/0009-0005-0221-3338

Po-Lan Su https://orcid.org/0000-0002-8470-6590

Chien-Chung Lin https://orcid.org/0000-0002-4739-5631

Supplemental material
Supplemental material for this article is available online.

References
1. Siegel RL, Miller KD and Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* 2020; 70(1): 7–30.
2. Miller KD, Siegel RL, Lin CC, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* 2016; 66(4): 271–289.
3. Raso MG and Wistuba II. Molecular pathogenesis of early-stage non-small cell lung cancer and a proposal for tissue banking to facilitate identification of new biomarkers. *J Thorac Oncol* 2007; 2(7, Suppl 3): S128–S135.
4. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. *N Engl J Med* 2016; 375(19): 1823–1833.
5. Baxevanos P and Mountzios G. Novel chemotherapy regimens for advanced lung cancer: have we reached a plateau? *Ann Transl Med* 2018; 6(8): 139.
6. Jeon DS, Kim HC, Kim SH, et al. Five-year overall survival and prognostic factors in patients with lung cancer: results from the Korean Association of Lung Cancer Registry (KALC-R) 2015. *Cancer Res Treat* 2023; 55(1): 103–111.
7. Howlader N, Noone AM, Krapcho M, et al. (eds.). *SEER Cancer Statistics Review, 1975–2011*. Bethesda, MD: National Cancer Institute, 2014.
8. Todaro GJ, Fryling C and De Larco JE. Transforming growth factors produced by certain human tumor cells: polypeptides that interact with epidermal growth factor receptors. *Proc Natl Acad Sci U S A* 1980; 77(9): 5258–5262.
9. Shi Y, Au JS, Thongprasert S, et al. A prospective, molecular epidemiology study of

EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J Thorac Oncol* 2014; 9(2): 154–162.

10. Dearden S, Stevens J, Wu YL, et al. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). *Ann Oncol* 2013; 24(9): 2371–2376.
11. Mok TS, Wu YL, Thongprasert S, et al. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009; 361(10): 947–957.
12. Zhou C, Wu YL, Chen G, et al. Final overall survival results from a randomised, phase III study of erlotinib versus chemotherapy as first-line treatment of EGFR mutation-positive advanced non-small-cell lung cancer (OPTIMAL, CTONG-0802). *Ann Oncol* 2015; 26(9): 1877–1883.
13. Wu YL, Zhou C, Hu CP, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014; 15(2): 213–222.
14. Juan O and Popat S. Treatment choice in epidermal growth factor receptor mutation-positive non-small cell lung carcinoma: latest evidence and clinical implications. *Ther Adv Med Oncol* 2017; 9(3): 201–216.
15. Shao SC, Chan YY, Kao Yang YH, et al. The Chang Gung Research Database—a multi-institutional electronic medical records database for real-world epidemiological studies in Taiwan. *Pharmacoepidemiol Drug Saf* 2019; 28(5): 593–600.
16. SEER*Explorer: An interactive website for SEER cancer statistics [Internet]. Surveillance Research Program, National Cancer Institute, https://seer.cancer.gov/statistics-network/explorer/. Data source(s): SEER Incidence Data, November 2023 Submission (1975-2021), SEER 22 registries (excluding Illinois and Massachusetts). Expected Survival Life Tables by Socio-Economic Standards. (2024, accessed 27 February 2025).
17. Lin JJ, Cardarella S, Lydon CA, et al. Five-year survival in EGFR-mutant metastatic lung adenocarcinoma treated with EGFR-TKIs. *J Thorac Oncol* 2016; 11(4): 556–565.
18. Kawaguchi T, Takada M, Kubo A, et al. Performance status and smoking status are independent favorable prognostic factors for survival in non-small cell lung cancer: a comprehensive analysis of 26,957 patients with NSCLC. *J Thorac Oncol* 2010; 5(5): 620–630.
19. Campos-Balea B, de Castro Carpeño J, Massutí B, et al. Prognostic factors for survival in patients with metastatic lung adenocarcinoma: an analysis of the SEER database. *Thorac Cancer* 2020; 11(11): 3357–3364.
20. Arcila ME, Nafa K, Chaft JE, et al. EGFR exon 20 insertion mutations in lung adenocarcinomas: prevalence, molecular heterogeneity, and clinicopathologic characteristics. *Mol Cancer Ther* 2013; 12(2): 220–229.
21. Kuijpers CCHJ, Hendriks LEL, Derks JL, et al. Association of molecular status and metastatic organs at diagnosis in patients with stage IV non-squamous non-small cell lung cancer. *Lung Cancer* 2018; 121: 76–81.
22. Leduc C, Merlio JP, Besse B, et al. Clinical and molecular characteristics of non-small-cell lung cancer (NSCLC) harboring EGFR mutation: results of the nationwide French Cooperative Thoracic Intergroup (IFCT) program. *Ann Oncol* 2017; 28(11): 2715–2724.
23. Kuan FC, Kuo LT, Chen MC, et al. Overall survival benefits of first-line EGFR tyrosine kinase inhibitors in EGFR-mutated non-small-cell lung cancers: a systematic review and meta-analysis. *Br J Cancer* 2015; 113(10): 1519–1528.
24. Zhang Y, Sheng J, Kang S, et al. Patients with exon 19 deletion were associated with longer progression-free survival compared to those with L858R mutation after first-line EGFR-TKIs for advanced non-small cell lung cancer: a meta-analysis. *PLoS One* 2014; 9(9): e107161.
25. Maemondo M, Inoue A, Kobayashi K, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010; 362(25): 2380–2388.
26. Asahina H, Yamazaki K, Kinoshita I, et al. A phase II trial of gefitinib as first-line therapy for advanced non-small cell lung cancer with epidermal growth factor receptor mutations. *Br J Cancer* 2006; 95(8): 998–1004.
27. Yang JJ, Zhou Q, Yan HH, et al. A phase III randomised controlled trial of erlotinib vs gefitinib in advanced non-small cell lung cancer with EGFR mutations. *Br J Cancer* 2017; 116(5): 568–574.
28. Park K, Tan EH, O'Byrne K, et al. Afatinib versus gefitinib as first-line treatment of patients with EGFR mutation-positive non-small-cell lung cancer (LUX-Lung 7): a phase 2B, open-label,

randomised controlled trial. *Lancet Oncol* 2016; 17(5): 577–589.

29. Soria JC, Felip E, Cobo M, et al. Afatinib versus erlotinib as second-line treatment of patients with advanced squamous cell carcinoma of the lung (LUX-Lung 8): an open-label randomised controlled phase 3 trial. *Lancet Oncol* 2015; 16(8): 897–907.
30. Sequist LV, Yang JC, Yamamoto N, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013; 31(27): 3327–3334.
31. Soria JC, Ohe Y, Vansteenkiste J, et al. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N Engl J Med* 2018; 378(2): 113–125.
32. Ramalingam SS, Vansteenkiste J, Planchard D, et al. Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. *N Engl J Med* 2020; 382(1): 41–50.
33. Lee SH, Menis J, Kim TM, et al. Pan-Asian adapted ESMO Clinical Practice Guidelines for the diagnosis, treatment and follow-up of patients with oncogene-addicted metastatic non-small-cell lung cancer. *ESMO Open* 2024; 9(12): 103996.

Visit Sage journals online journals.sagepub.com/home/tam

Sage journals